undated
probably ca. 1981

Genetics and the Synthesis of Proteins

N. H. Horowitz

Biology Division, California Institute of Technology

Pasadena, California 91125

Symposium on The Origins of Modern Biochemistry: A Retrospect on Proteins

Genes, Proteins, and the Genetic System

The discovery and elucidation of the connection between genes and proteins is one of the major accomplishments of 20th century science. Nothing in biology illuminates more clearly the fundamental organization of living systems than does the gene-protein relationship. We see that every organism has a genetic heritage that consists apparently entirely of specifications for the synthesis of an array of protein molecules, including their structure, time of appearance, and rate of production. From this initial input, all other aspects of the organism follow--its structure, development, metabolism, and behavior--insofar as these are genetically determined. The genetic specifications are an evolutionary product, generated by random mutations in DNA and screened by natural selection. They are a record of discovered solutions to the problems of survival encountered by the species during its long history. Without this historical record, life could not exist, because survival depends on the ability of the organism to synthesize a large variety of proteins, but proteins are highly improbable structures. If every generation had to discover for itself how to assemble amino acids in the correct sequences to produce useful proteins, survival would be impossible. Hence the need to preserve and transmit sequence information from generation to generation.

Until about 30 years ago, it was generally believed that proteins themselves perform this genetic function. Avery and those who followed him showed the error of this belief. Proteins cannot serve as templates for their own replication (at least they cannot do so very effectively) and therefore they cannot function as genes. Nucleic acids, however, can serve as their own templates, and hence they are suited to be carriers of hereditary information, but they cannot perform the many catalytic functions that are essential for the life of the cell and that proteins perform so readily. Hence the dual system, nucleic acids and proteins, interlocking and interdependent, each indispensable for the existence of the other. The properties of this remarkable system--the "genetic system"--coincide with the essential properties of living matter; that is to say, it is capable of duplicating itself, of mutating and duplicating its mutations, and of evolving adaptively.

Garrod's Discovery

The first intimation that biological inheritance is concerned with the synthesis of proteins came shortly after the rediscovery of the laws of Mendel. In 1902, just two years after the rediscovery, Archibald Garrod[1] published a paper in which he suggested that alcaptonuria in man is inherited as a simple Mendelian recessive. In this suggestion he had the

support of William Bateson, the leading British proponent of Mendelism. By 1909, Garrod's investigations had brought him to the further conclusion that alcaptonuria results from the lack of an enzyme, present in normal individuals, that opens the ring of homogentisic acid. This he published in his classic treatise, "Inborn Errors of Metabolism," along with tentative evidence suggesting that several other inherited defects in man have a similar basis.[2]

Garrod's discovery, made practically at the outset of modern genetics, suffered a familiar fate: it was ignored and forgotten by geneticists for over 30 years, when the same principles were rediscovered in Neurospora by Beadle and Tatum. Although Garrod is now honored as the father of biochemical genetics, the fact is that his work had no influence on the development of genetics, which would have been the same had he never lived. The same thing, of course, is true of Mendel. Garrod's fate differed from Mendel's in one important respect, however: his work on alcaptonuria was highly regarded by biochemists and was accepted into the body of biochemical knowledge intact. It was discussed in a biochemistry course I took in 1935. What interested biochemists, however, were Garrod's findings in connection with the metabolism of phenylalanine and tyrosine; the possible implications of these findings for the nature of gene action completely escaped them.

Geneticists, too, had other, more urgent concerns in the early decades of the century. First was the question of the validity and

generality of Mendelism as a description of heredity in plants and animals. Then came the chromosome theory, the proof of this theory, and the demonstration that genes can be mapped on the chromosomes. It was not until these matters had been settled that the question of how genes produce their effects became pressing.

In his book on the history of genetics, A. H. Sturtevant[3] gives two reasons for the failure of geneticists to appreciate Garrod's discovery. At first, they did not understand what Garrod was talking about because they were ignorant of biochemistry. Later, when they were more literate in biochemistry, they understood him but did not accept his findings as generally applicable; they were convinced that development was too complex to be explained by any simple theory of gene action. The view held by most geneticists until the 1950s was that genes are manifold, or pleiotropic, in their action. This view precluded serious consideration of Garrod's findings, and later it delayed acceptance of the extensive and convincing evidence obtained in Neurospora for a one-to-one relation between genes and enzymes. According to Sturtevant, the notion of pleiotropy arose from the early studies of De Vries on the so-called mutations in Oenothera Lamarckiana, the evening primrose. These "mutations," which for a long time defied analysis, are now known not to be single-gene events, but multiple genetic changes resulting from recombination within the unusually complex chromosomes of Oenothera.

In a conversation I had with him a few years before he died, Sturtevant told me that a theoretical argument by E. B. Wilson had also been important in obscuring the significance of Garrod's findings. I believe that the argument Sturtevant referred to was the following from the 3rd edition of The Cell [4]:

> "In what sense can the chromosomes be considered as agents of determination? By many writers they have been treated as the actual and even as the exclusive 'bearers of heredity' . . . Many writers, while avoiding this particular usage, have referred to the chromosomes, or their components as 'determiners' of corresponding characters; but this term, too, is becoming obsolete save as a convenient descriptive device. The whole tendency of modern investigation has been towards a different and more rational conception which recognizes the fact that the egg is a reaction-system and that (to cite an earlier statement) 'the whole germinal complex is directly or indirectly involved in the production of every character.' Genetic research is constantly bringing to light new cases of the cooperation of several or many factors in the production of single characters; and it is possible that all the chromosomes, or even all of the units which they contain, may be concerned in the production of every character."*

*A. D. Hershey[34] quotes another, very similar, statement by E. B. Wilson in an interesting essay on the state of genetics written in 1970.

The view expressed in this quotation is one that was widespread among geneticists in the '30s and '40s. In a sense, this view is perfectly correct; but in another, and equally valid sense, it is totally wrong. It was a long time before these different ways of regarding gene action could be sorted out. In the meantime, Garrod's important discovery was forgotten.

Neurospora and "One Gene-One Enzyme"

In the decades following the publication of Inborn Errors a number of starts were made on the biochemical analysis of mutant phenotypes in a variety of plants and animals, but the subject was not essentially advanced beyond the point where Garrod had left it until 1941, when Beadle and Tatum described the first nutritional mutants (also called "auxotrophs") in Neurospora. (The initial experiments used both N. sitophila and N. crassa, but all subsequent work was with N. crassa.) In the interval, genetics had developed in almost total isolation from biochemistry and, in fact, from all physical sciences. This isolation was described in the following way by Sturtevant and Beadle in the preface to their "Introduction to Genetics," published in 1939[6]:

> "Physics, chemistry, astronomy, and physiology all deal with atoms, molecules, electrons, centimeters, seconds, grams--their measuring systems are all reducible to these common units. Genetics has none of these as a recognizable component

in its fundamental units, yet it is a mathematically formulated subject that is logically complete and self-contained."

The 1941 paper of Beadle and Tatum marks the end of this isolation of genetics from the physical sciences. The recovery of single-gene mutants in which specific biosynthetic pathways were blocked opened a new dimension in the study of gene action. In place of the chemically undefined morphological mutations that up to that time had formed the main working material of genetics, there was now a wealth of inherited metabolic defects which were comprehensible in terms of known biochemistry. Unlike alcaptonuria in man, the Neurospora results could not be explained away as a singularity. Furthermore, the methods devised by Beadle and Tatum were applicable to other microorganisms, as Tatum soon showed. He was able to induce the same kinds of mutations in Escherichia coli (strain K-12, by lucky chance).[7] These mutants were later used by Lederberg and Tatum[8] to demonstrate sexual recombination in E. coli, itself a major event in the history of genetics.

Study of the Neurospora mutants soon made it clear that at the level of metabolic reactions genes are not pleiotropic at all, but are highly restricted in their range of action. Just as in alcaptonuria, the "biochemical" mutants of Neurospora, as they were then called, were blocked in single steps of metabolism.[9] The seeming exceptions turned out to be exceptions that proved the rule. For example, a single-gene mutant

that required two amino acids--methionine and threonine--was found to be blocked not in two pathways, but in the synthesis of homoserine, a previously unrecognized common precursor. Again, a mutant that required both isoleucine and valine for growth was, after much investigation, shown actually to be blocked in two pathways, but these pathways--the last few steps in the synthesis of these two amino acids--are catalysed by the same enzymes.[10]

The Neurospora results were summarized in the "one gene-one enzyme" hypothesis--i.e., a given gene is involved in the synthesis of a single enzyme or other protein. When the gene mutates, the enzyme is defective or is simply not made. As Beadle conceived it in 1945, the gene acted as a "master molecule or templet in directing the final configuration of the protein molecule as it is put together from its component parts."[11] This theory was advanced before any proof of enzyme involvement had been obtained and of course before anything was known about the chemical nature of the gene. It was an inference based on analysis of the growth requirements of the mutants, on their mode of inheritance, and on the identification of metabolic intermediates that accumulated in blocked pathways. The first direct demonstration of an enzymatic deficiency in a Neurospora mutant was of tryptophan synthetase in a tryptophan-requiring mutant, by Mitchell and Lein.[12] Many other examples followed.

Beadle's theory was greeted with hostility. It seemed that, despite the evidence, many geneticists preferred to believe Wilson's doctrine that

every gene is concerned in the production of every character. The one gene-one enzyme hypothesis was denounced as unverifiable and also unfalsifiable. It was alleged to be based on a selection procedure that insured that only mutations supporting the theory would be detected. It was criticized for being too simple to explain all of the complexities of metabolism. Critiques published at the time are but pale shadows of the unpublished objections that were voiced in the '40s and '50s at the Cold Spring Harbor symposia and wherever else geneticists gathered.

The debate lasted on and off for years; it finally ended with the vindication of "one gene-one enzyme." We can date this milestone with the demonstration, simultaneously in Yanofsky's and Brenner's laboratories, of the colinearity of gene and protein in E. coli and bacteriophage, respectively.[13,14] Yanofsky and his coworkers mapped a series of missense mutations in the structural gene for the A polypeptide of tryptophan synthetase of E. coli, and at the same time they mapped the corresponding amino acid replacements in the polypeptide. The two maps were superimposable within the limits of error of the measurements. The Brenner group mapped a series of nonsense mutations in the gene encoding the head protein of phage T4 and showed that the locations of the resulting interruptions in the elongation of the polypeptide were colinear with the genetic map. Colinearity is a sufficient condition for "one gene-one enzyme," although not a necessary one. In addition, since it was already becoming clear

that the active configuration of proteins is determined by their amino acid sequence,[15] it followed that the gene carries all of the unique information needed to specify the enzyme. One gene-one enzyme was just a special case of the general relation: one gene-one polypeptide.

By 1964, when the colinearity papers were published, the terms of the discussion had changed considerably from what they had been in 1945 when Beadle advanced the notion of a simple relation between genes and enzymes. The chemical structure of the gene had by then been discovered, and great advances had been made in unravelling the mechanism of protein synthesis. It was no longer necessary to infer the nature of gene action from the results of genetic experiments, since gene action was rapidly becoming amenable to direct study at the molecular level. While the colinearity experiments of Yanofsky and of Brenner still contained recognizably classical features, they went far beyond classical genetics in their analysis of both the genotype and, especially, the phenotype. As will be seen later, we have reached the point today where the gene as well as its products are analyzable into their ultimate structural units, and new discoveries are being made at this level.

Sickle-cell hemoglobin

In the course of the one gene-one enzyme debate, a number of important results were obtained that should be mentioned here. Beadle's use of the word "templet" to describe the role of the gene suggests a

mechanism of protein synthesis that is remote from the actual mechanism as we know it today, but his point, obviously, was not to propose a mechanism of protein synthesis, but to suggest that the gene determines the specific properties of the enzyme, not just its presence or absence. That this is so was first demonstrated not in Neurospora or E. coli, but in man. In 1949, Pauling and coworkers[16] showed that sickle-cell hemoglobin has a higher isoelectric point than normal hemoglobin, the difference amounting to 2-4 net charges per molecule. In 1957, Ingram[17] found that sickle-cell hemoglobin has valine in place of glutamic acid in position 6 of the β chains. This was the first demonstration that amino acid substitution can result from gene mutation.

Temperature-sensitive mutants

A class of mutants that was especially useful in establishing the relation between genes and proteins was the temperature-sensitive, or temperature-conditional, class. These mutants show their genetic defect at particular temperatures--usually above 30° C--while they are normal, or nearly so, at other--usually lower--temperatures. They are caused, as we know now, by mutations in the structural gene for an enzyme that result in the production of thermolabile forms of the enzyme. When they were first discovered, however, it was not known whether it was the enzyme or the enzyme-synthesizing mechanism that had become thermolabile.

Temperature-conditional mutants were first found in the course of the mutant hunt that ran more or less continuously in Beadle's laboratory at Stanford in the '40s. The decision to search for such mutants had been made following publication of a paper by Stokes et al.[18] showing that one of the three original mutants of Beadle and Tatum, a pyridoxin-requiring strain, was pH-sensitive: its need for pyridoxin was displayed only when the pH of the medium was below 5.8. Above pH 5.8, it synthesized the vitamin. This finding suggested that a similar class of temperature-sensitive mutants might exist, and this was soon confirmed.[19, 20]

These mutants were especially valuable because they made it possible to detect, in the form of temperature-conditional alleles, genes whose ordinary mutations would be lethal and unrecoverable. This property made it possible to answer a fundamental criticism of the one gene-one enzyme hypothesis that had been advanced by Max Delbrück; namely, that the method of detecting nutritional mutants was such that it was inherently unlikely that any mutants with complex nutritional requirements would be recovered. If this criticism was valid, and if mutants with multiple metabolic defects were a significant fraction of the total, then this should be revealed by analysis of temperature-conditional mutants, which are selected only for their failure to grow at certain temperatures. Specifically, by placing the mutants at the temperature at which their mutant character is manifested, it could be determined what fraction of

them failed to grow when supplied with the standard complete medium used in the standard selection procedure. When I applied this test to the known temperature-conditional mutants of Neurospora in 1950, I could find little evidence for selection of the kind Delbrück had postulated.[9] Leupold and I then examined a much larger number of temperature-conditional mutants in E. coli and found even less evidence for selection against multifunctional losses than in Neurospora.[21] This result was very reassuring for the one gene-one enzyme theory.

In 1952, Maas and Davis[22] described a temperature-conditional mutant of E. coli that required exogenous pantothenic acid at temperatures above 30°C, but not at 25°C. They were able to show that the mutant produces a thermolabile form of pantothenate synthetase, the enzyme that couples pantoic acid and β-alanine. This was the first evidence that temperature-sensitive mutations affect the enzyme, not the enzyme-synthesizing apparatus. A little later, Fling and I found a gene in Neurospora that determines both the thermostability[23] and the electrophoretic mobility[24] of the enzyme tyrosinase.

Occasionally, mutants were found with reversed temperature sensitivity--i.e., they were phenotypically mutant at low temperatures but normal at high temperatures. These were more difficult to account for, since no models were known to us of enzymes that were inactivated by lowering the temperature by a few degrees. In 1957, however, Fincham[25] found that a glutamic acid-requiring mutant of Neurospora with reversed temperature sensitivity produced a glutamic dehydrogenase with precisely

the properties needed to explain the phenotype; that is, the enzyme is active at temperatures above 25°C, but is inactivated reversibly at 20°C. All of this strengthened the idea that the structure of proteins is genetically determined.

Recent Developments

The foregoing is a brief summary of the investigations--with emphasis on those I had some personal involvement in--that led to the picture of the gene-protein relationship that has been accepted for nearly two decades. According to this picture, a continuous segment of DNA, the gene, codes for a unique polypeptide which is a linear representation of the DNA. Until very recently, this picture was thought to be applicable throughout the living world. In the past year, however, findings made possible by powerful new methods for amplifying and sequencing nucleic acids have shown that the accepted model is far from universally applicable.

The new findings are of two kinds. First, several examples are now known of gene overlaps--i.e., stretches of DNA that are shared by two or even three genes.[26-28] A shared sequence may be translated in the same reading frame for two proteins, in which case the proteins have amino acid sequences in common, or it may be translated out of phase. In the latter case, one DNA sequence can have three different translations. These contradictions of the one gene-one polypeptide rule have been found so far only

in viruses, where information compression presumably has a strong selective advantage.

The second kind of unexpected finding violates the colinearity rule. In a growing number of cases, eucaryotic structural genes are being found to contain interpolated DNA sequences that are not represented in the mRNA or tRNA that is read from these genes.[29-33] The inserts are known to be transcribed in at least some cases (and are presumed to be transcribed in all cases), but they are eliminated in the processing of the transcript. The significance of the interpolated DNA is unknown at this time, although there are some plausible suggestions.

It seems likely that these contradictions of the, until now, accepted model of the relationship between genes and proteins will turn out to be special evolutionary adaptations for life as a virus or as a eucaryote--that is, higher order refinements of the simple basic pattern. In any case, these fascinating results are doubtless just the forerunners of discoveries that the new techniques now available will make possible. It is clear that the story of genes and proteins is not yet over.

References

1. GARROD, A. E. 1902. Lancet 1902 (ii): 1616-1620.
2. GARROD, A. E. 1909. Inborn Errors of Metabolism. Frowde; Hodder & Stoughton. London.
3. STURTEVANT, A. H. 1965. A History of Genetics. Harper & Row. New York.
4. WILSON, E. B. 1925. The Cell in Development and Heredity. 3rd edit. Macmillan. New York. pp. 975-976.
5. BEADLE, G. W. & E. L. TATUM. 1941. Proc. Natl. Acad. Sci. U. S. 27: 499-506.
6. STURTEVANT, A. H. & G. W. BEADLE. 1939. An Introduction to Genetics. W. B. Saunders Co. Philadelphia.
7. GRAY, C. H. & E. L. TATUM. 1944. Proc. Natl. Acad. Sci. U. S. 30: 404-410.
8. LEDERBERG, J. & E. L. TATUM. 1946. Cold Spring Harbor Symp. Quant. Biol. 11: 113-114.
9. HOROWITZ, N. H. 1950. Advances in Genetics 3: 33-71.
10. MYERS, J. W. & E. A. ADELBERG. 1954. Proc. Natl. Acad. Sci. U. S. 40: 493-499.
11. BEADLE, G. W. 1945. Chem. Revs. 37: 15-96.
12. MITCHELL, H. K. & J. LEIN. 1948. J. Biol. Chem. 175: 481-482.
13. YANOFSKY, C., B. C. CARLTON, J. R. GUEST, D. R. HELINSKI & U. HENNING. 1964. Proc. Natl. Acad. Sci. U. S. 51: 266-272.

14. SARABHAI, A., A. O. W. STRETTON, S. BRENNER, & A. BOLLE. 1964. Nature 201: 13-17.

15. ANFINSEN, C. B. 1973. Science 181: 223-230.

16. PAULING, L., H. A. ITANO, S. J. SINGER & I. C. WELLS. 1949. Science 110: 543-548.

17. INGRAM, V. 1957. Nature 180: 326-328.

18. STOKES, J. L., J. W. FOSTER & C. R. WOODWARD, JR. 1943. Arch. Biochem. 2: 235-245.

19. MITCHELL, H. K. & M. B. HOULAHAN. 1946. Fed. Proc. 3: 370-375.

20. MITCHELL, H. K. & M. B. HOULAHAN. 1946. Amer. J. Bot. 33: 31-35.

21. HOROWITZ, N. H. & U. LEUPOLD. 1951. Cold Spring Harbor Symp. Quant. Biol. 16: 56-65.

22. MAAS, W. K. & B. D. DAVIS. 1952. Proc. Natl. Acad. Sci. U. S. 38: 785-797.

23. HOROWITZ, N. H. & M. FLING. 1953. Genetics 38: 360-374.

24. HOROWITZ, N. H., M. FLING, H. MACLEOD, & N. SUEOKA. 1961. Genetics 46: 1015-1024.

25. FINCHAM, J. R. S. 1957. Biochem. J. 65: 721-728.

26. SANGER, F., G. M. AIR, B. G. BARRELL, N. L. BROWN, A. R. COULSON, J. C. FIDDES, C. A. HUTCHISON III, P. M. SLOCOMBE & M. SMITH. 1977. Nature 687-695.

27. SHAW, D. C., J. E. WALKER, F. D. NORTHROP, B. G. BARELL, G. N. GODSON & J. C. FIDDES. 1978. Nature 272: 510-515.

28. CRAWFORD, L. V., C. N. COLE, A. E. SMITH, E. PAUCHA, P. TEGTMEYER, K. RUNDELL & P. BERG. 1978. Proc. Natl. Acad. Sci. U. S. 75: 117-121.

29. GOODMAN, H. M., M. V. OLSON & B. D. HALL. 1977. Proc. Natl. Acad. Sci. U. S. 74: 5453-5457.

30. VALENZUELA, P., A. VENEGAS, F. WEINBERG, R. BISHOP & W. J. RUTTER. 1978. Proc. Natl. Acad. Sci. U. S. 75: 190-194.

31. TILGHMAN, S. M., D. C. TEIMEIER, J. G. SEIDMAN, B. M. PETERLIN, M. SULLIVAN, J. V. MAIZEL & P. LEDER. 1978. Proc. Natl. Acad. Sci. 75: 725-729.

32. TILGHMAN, S. M., P. J. CURTIS, D. C. TEIMEIER, P. LEDER & C. WEISSMANN. 1978. Proc. Natl. Acad. Sci. U. S. 75: 1309-1313.

33. WEINSTOCK, R., R. SWEET, M. WEISS, H. CEDAR & R. AXEL. 1978. Proc. Natl. Acad. Sci. 75: 1299-1303.

34. HERSHEY, A. D. 1970. Nature 226: 697-700.